

Session 2d - 2

# TAMU Computational Vision for Real-Time Fusion of Stereo & LADAR Data

*James Turner & John Junkins*

Texas A&M University (TAMU)

*Brad Sallee*

System Processes Engineering Corporation (SPEC)

15-16 July 2014

**NASA In Space Inspection Workshop**

*We are pleased to acknowledge our stellar collaborators:*

*R. Noster (SPEC), M. Manoranjan (BU), et al,*

# Agenda

- Overview of Real Time Proximity Mapping Challenges and Novel Approaches to Meet Them
  - Essential Ideas & Experimental Demonstrations (LASR Lab.)
  - TAMU Computational Vision Pipeline
- HD6D LADAR, Vision Sensors and Data Fusion
  - LADAR Innovations
  - HD6D: An Embedded System for Fusion of LADAR and HD Image Data for Simultaneous Location and Mapping
- Accomplishments and Directions
- Concluding Remarks



# HD6D

## Real-Time Vision-Based Mapping

*“real-time fusion of ultra high speed scanning LADAR, and high definition imaging to enable orders of magnitude improvement of proximity navigation and mapping”*

J. Junkins, et al  
Texas A&M University

Brad Salley, et al  
SPEC

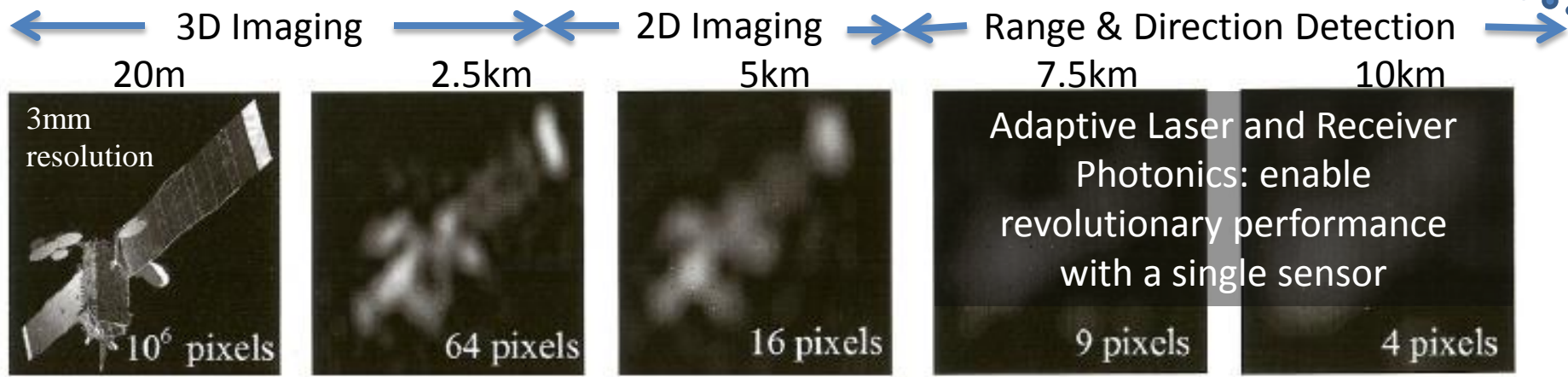
March 28, 2011

# Notional Performance of HD6D LADAR

Revolutionary HD Sensor Designed for Wide Range of Space Protection Applications

**Enables Wide Variation in Standoff Range from Sensor to Object**  
**Provides HD Range Image @ Video Rates**

Current Design:  
30 Km Max



Note, due to the foveal scan intrinsic to the HD3D Risley Scanning Prism optical design, the spatial sampling is over 2x as dense near the center of the FOV, and as consequence, the pixels over-sample the object at long range. This has the consequence of a degree of super-resolution → very advantageous when there are fewer than 100 pixels on the object.

Small SWAP 3D-Lensless HD LADAR, Fused with Imagery for Unprecedented Real-Time Estimation of Sensor-to-Imaged Vehicle Pose & HD Geometry and Surface Characteristics

Proprietary Information of Texas A&M University and System Processes and Engineering Corporation

POCs: [junkins@tamu.edu](mailto:junkins@tamu.edu), [turner@tamu.edu](mailto:turner@tamu.edu), [noster@spec.com](mailto:noster@spec.com), [sallee@spec.com](mailto:sallee@spec.com)

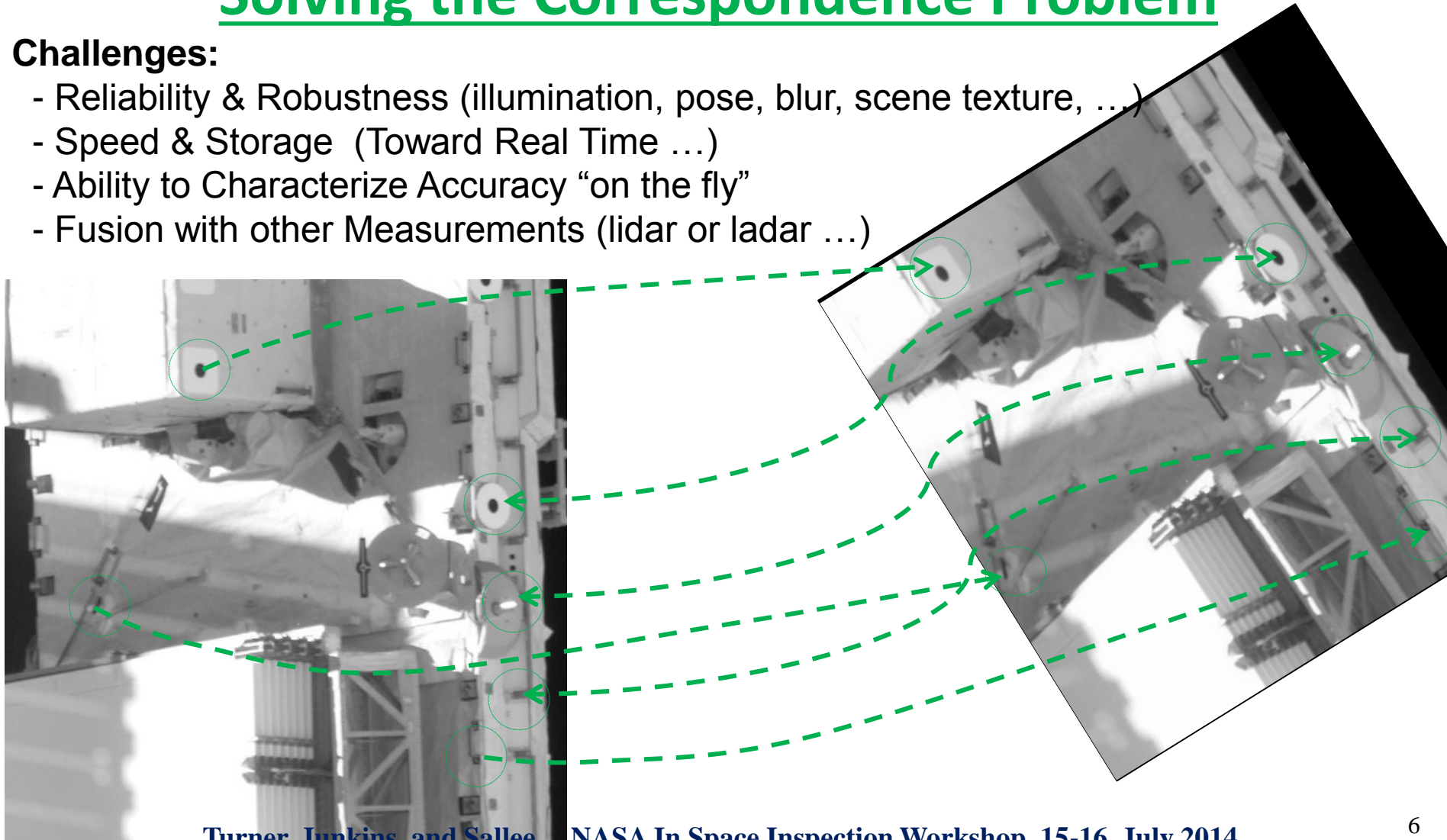




# The Crucial Step in 3D Mapping From Overlapping Images is Solving the Correspondence Problem

## Challenges:

- Reliability & Robustness (illumination, pose, blur, scene texture, ...)
- Speed & Storage (Toward Real Time ...)
- Ability to Characterize Accuracy “on the fly”
- Fusion with other Measurements (lidar or ladar ...)



- **Feature Detection/Characterization/Matching:**

1. SIFT ([David Lowe](#) ,1999-, Difference of Gaussian Blurs)
2. SURF([Herbert Bay](#) ,et al. 2006-, uses Haar Wavelet Operator)
3. SVT(Chen, et al, 2008-, Multi-view logic tree descriptor)
4. **Randomized Trees** (Lepetit, et al, 2005-)
5. **MVKP** (Wang et al, 2006-, Walsh-Hadamard kernels)
6. **BRIEF** (Calonder, et al, Binary Robust Independent Elementary Features. 2008-)

*All of these strive to efficiently and robustly locate the most distinct features from texture maps and characterize them uniquely; tolerating blur, distortion, large scale, linear & angular displacement changes*

### Assessments Based on Our Research:

- Reliability favors Randomized Trees, BRIEF, SIFT, SURF
- Speed favors 2, 4-6, however, all 6 operate in a GPU or FPGA @ frame rate ...
- All methods fail reliably for sufficiently difficult scenes, all work well for sufficiently textured, nicely illuminated, and near-planar scenes
- None of these consider 3D Geometry simultaneously with Texture => *Flewelling R&D*
- Pervasive Problem, Characterizing “How Well” these methods are working *on the fly!*
- Any of these can serve more than satisfactorily to find the features and form the initial hypothesis in our vision pipeline (will justify) => statistical rejection of spurious features.
- Other Methods are more attractive for recursive feature tracking after initial detection.

## ***BOTTOM LINE UP FRONT:***

***Stereo Vision Alone*** is not an adequate basis for solving the family of SLAM problems of interest, because

- Scene, Lighting, Geometry, and Range-to-Object variations causes failures that are difficult to anticipate (and/or tolerate!)

## ***WHAT IS NEEDED:***

- **New Sensor Technologies** to Fuse with Stereo Vision, to render the SLAM process far less susceptible to the above problems.
- **Algorithms that make use of the Feature Point redundancy**, => better tolerate the wide variability of lighting, texture, and image quality => **statistically select the best features** and **characterize accuracy, “ON THE FLY”**.

## ***WE ARE RESEARCHING & BUILDING $\beta$ PROTOTYPES OF:***

- ***A revolutionary sensor: HD6D***, a new generation LADAR  $\sim 12 \times 10^6$  range measurements/sec  $\pm 3\text{mm } 1\sigma$ , co-registered & synced with a HD video camera.
- ***A novel software pipeline: fuse overlapping LADAR and HD imagery in real time***
  - Basic Advance: Point Cloud Fusion Formulation => ***Rigorous Linearization***
  - HP\_RANSAC: ***statistical characterization & spurious feature rejection***



$$\begin{aligned} r'_i &= Rr_i + \tau, & \Leftrightarrow & \begin{bmatrix} x'_i \\ y'_i \\ z'_i \end{bmatrix} = R \begin{bmatrix} x_i \\ y_i \\ z_i \end{bmatrix} + \begin{bmatrix} \tau_x \\ \tau_y \\ \tau_z \end{bmatrix} \Rightarrow \left\{ \text{For Registration:} \right. \\ r'_c &= Rr_c + \tau & & \left. \{R(q), \tau\} = ?? \right\} \end{aligned}$$

$$z_i = r_i - r_c, \quad \boxed{z'_i = r'_i - r'_c = Rz_i}, \quad R = \overbrace{R(q)}^{\text{Cayley Transform}} = [I + \tilde{q}]^{-1}[I - \tilde{q}], \quad q = \tan \frac{\phi}{2} \hat{e}$$

Re-arrange as a Rigorous Linearization:

$$z'_i = Rz_i = [I + \tilde{q}]^{-1}[I - \tilde{q}]z_i \Rightarrow \begin{cases} [I + \tilde{q}]z'_i = [I - \tilde{q}]z_i \\ -[\tilde{q}]\{z_i + z'_i\} = \{z_i - z'_i\} \\ \boxed{[\tilde{s}_i]q = d_i}, \quad s_i = \{z_i + z'_i\}, \quad d_i = \{z_i - z'_i\} \end{cases}$$

Then the best (least square) rotation and translation estimates are obtained w/o iteration:

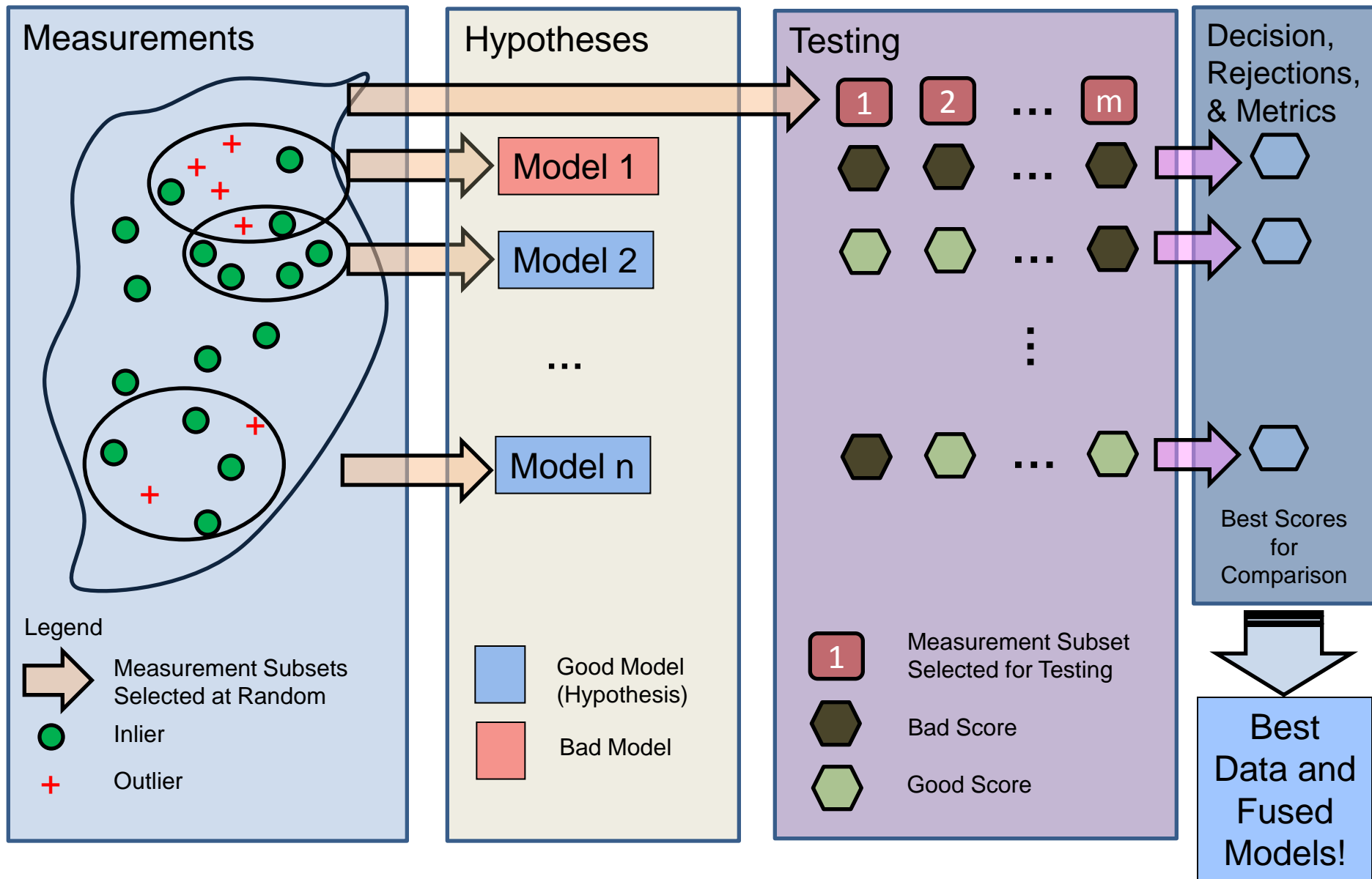
$$\hat{q} = [H^T H]^{-1} H^T d, \quad \hat{\tau} = \frac{1}{N} \sum_{i=1}^N \{r'_i - R(\hat{q})r_i\}, \quad \text{where: } d = \begin{Bmatrix} d_1 \\ \vdots \\ d_N \end{Bmatrix}, \quad H = \begin{Bmatrix} [\tilde{s}_1] \\ \vdots \\ [\tilde{s}_N] \end{Bmatrix}$$

$$v = \begin{Bmatrix} v_x \\ v_y \\ v_z \end{Bmatrix}, \quad \tilde{v} \equiv \begin{bmatrix} 0 & -v_z & v_y \\ v_z & 0 & -v_x \\ -v_y & v_x & 0 \end{bmatrix}$$

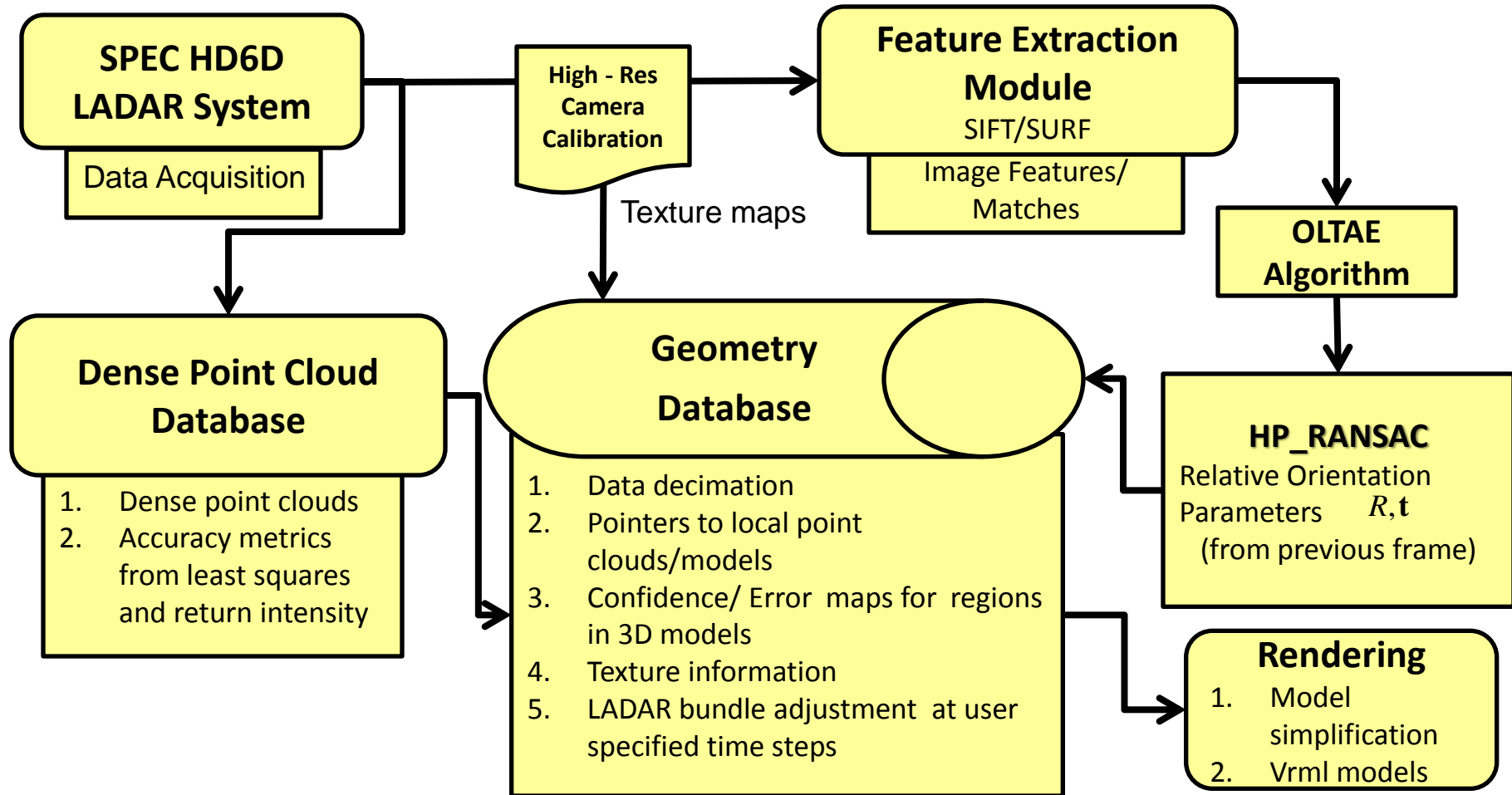
Residual error statistics:

$\epsilon_i = r'_i - \{[R(\hat{q})]r_i + \hat{\tau}\} = \text{ith residual vector, } i = 1, N, \text{ typically, } |\epsilon_i| < 4[\sigma_x^2 + \sigma_y^2 + \sigma_z^2]^{1/2}$

$$\text{mean: } \mu = \frac{1}{N} \sum_{i=1}^N \epsilon_i \sim 0, \quad \text{covariance: } \begin{bmatrix} \sigma_x^2 & \rho_{xy}\sigma_x\sigma_y & \rho_{xz}\sigma_x\sigma_z \\ & \sigma_y^2 & \rho_{yz}\sigma_y\sigma_z \\ \text{sym} & & \sigma_z^2 \end{bmatrix} = \frac{1}{N} \sum_{i=1}^N \epsilon_i \epsilon_i^T$$



## Optimal Computational Vision Pipeline (OCVP): Data Flow



**All modules in place (as can be seen from the Kinect demo)  
Integration with SPEC/TAMU HD6D System starts next month**



## PRODUCT SPECIFICATION

Property	PrimeSensor Spec
Field of View (Horizontal, Vertical, Diagonal)	58° H, 40° V, 70° D
Depth image size	VGA (640x480)
Spatial x/y resolution (@2m distance from sensor)	3mm
<u>Depth z resolution</u> (@2m distance from sensor)	<u>1cm</u>
<u>Maximal image throughput</u> (frame rate)	<u>60fps</u>
Average image latency in full VGA resolution	40msec
<u>Operation range</u>	<u>0.8m - 3.5m</u>

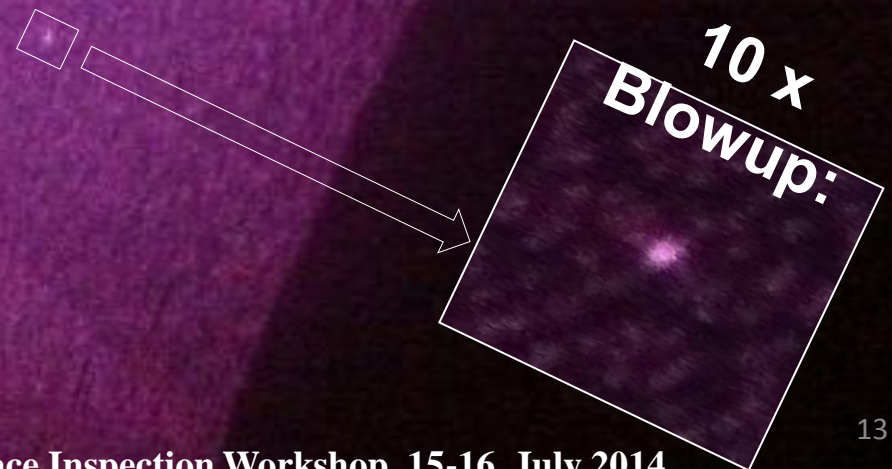
Property	PrimeSensor Spec
<u>Color image size</u>	<u>UXGA (1600x1200)</u>
Audio: built-in microphones	2 mics
Audio: digital inputs	4 inputs
Data interface	USB 2.0
Power supply	USB 2.0
<u>Power consumption</u>	<u>2.25W</u>
Dimensions (Width x Height x Depth)	14cm x 3.5cm x 5cm
Operation environment (every lighting condition)	indoor
Operating temperature	0°C - 40°C

Current Kinect sensor configuration applicable to the “final 4m”

=> Concept extendable with moderate design revisions ~20m range with “reasonable SWAP”

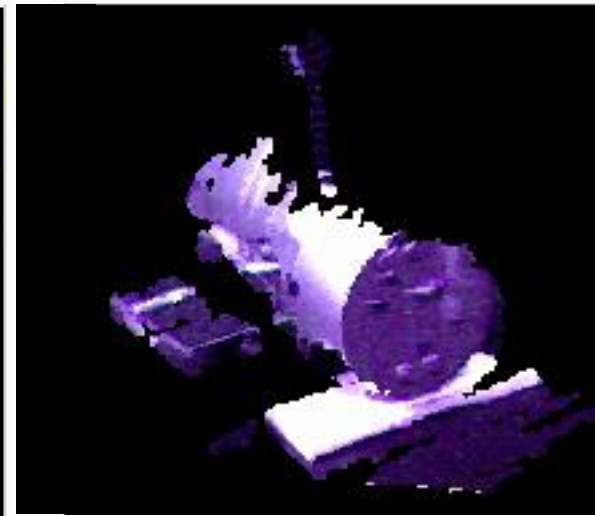
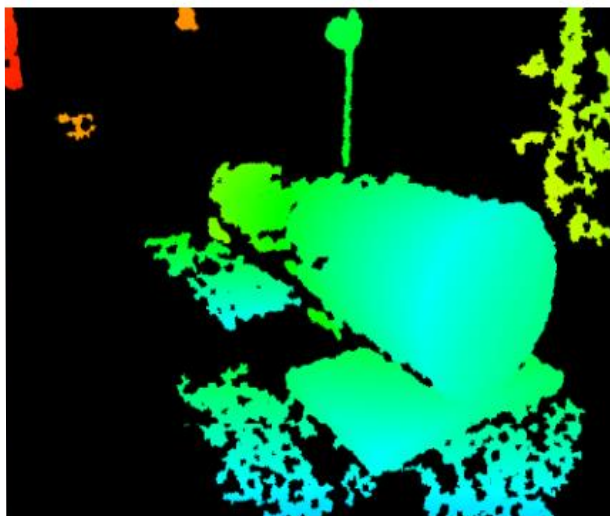
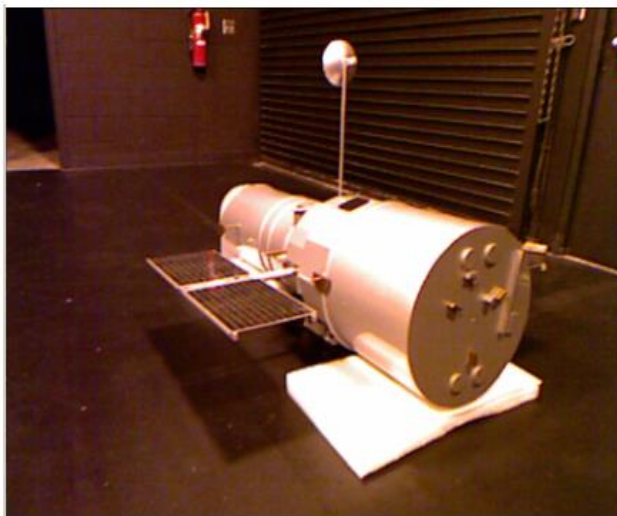


# Structure of Kinect HD IR Projected Light: Nine Intensity Zones, Nine Bright IR Targets, Unique Quasi-Random Fainter Texture

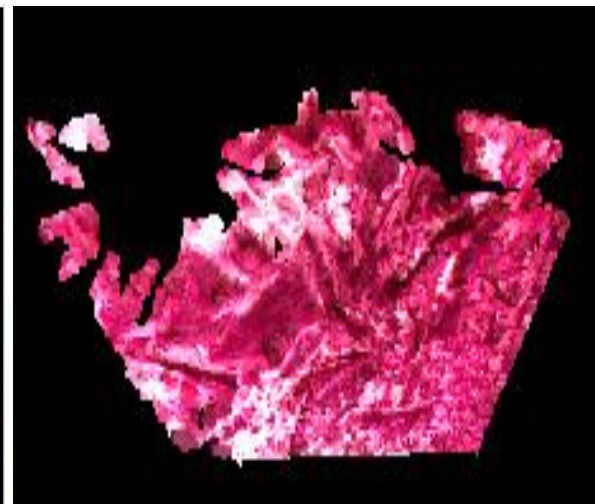
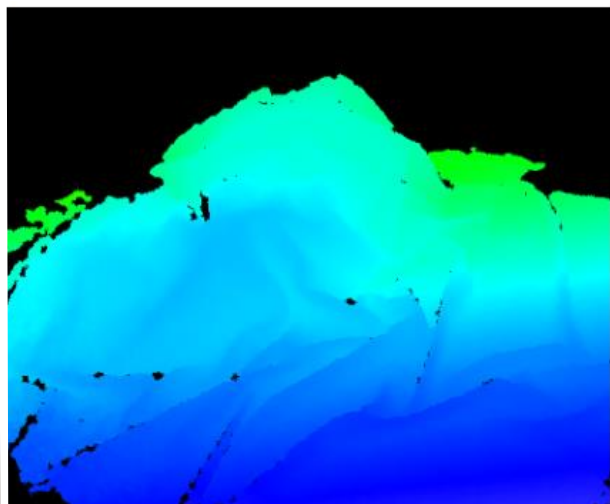


# “One-Frame” 3D Reconstructions From Standard Kinect Sensor

10% Scale Model of Hubble (smooth ~textureless model)



Highly Textured & Occcluded Artificial Terrain (“laboratory Afghanistan”)





# Real Time Relative Navigation and Scene Mapping From Optical Measurements

## Challenges:

- **Current scanning LIDAR/LADARs are too slow to fuse efficiently with HD video**
  - *HD Video frame rates are  $>12\text{HZ}$*
  - *Conventional LIDAR/LADAR HD point clouds (# of range measurements =  $\sim$  # of pixels =  $\sim 10^6$ ) are captured up to three orders of magnitude slower than video rates (scanning lidar typically collects range measurements at  $\sim 10\text{KZ}$ , to be compatible with HD video, need to collect range measurements at  $\sim 12\text{MHZ}$ )*
    - *The motion of the lidar sensor while measuring the scene cannot be neglected and therefore solving the **correspondence problem** between the lidar range measurements and the “instantaneous” HD camera image introduces a computational bottleneck, typically not compatible with real time fusion*
- **While FLASH LIDAR and staring LIDAR technologies are advancing and provide the frame rate to solve the above problem, however**
  - Pixel formats have not yet exceeded  $512 \times 512$  pixels, and most are either  $128 \times 128$  or  $256 \times 256$  pixels.
  - Also, the radially diffuse nature of flash lidar means that it is unsuitable for long standoff ranges
- **SWAP Challenge: Current sensor technologies for vision-based mapping and scanning LIDAR are too large for many Size–Weight–Power constrained missions.**
- **Question: Can the next generation LADAR enjoy the advantages of both scanning lidar and flash lidar? *Enable Real Time Fusion of HD LADAR and Video?***

# We are at the Confluence of three sets of Rapidly Advancing Technologies

- **Advances in Computing**
  - Multi-Teraflop Personal (Cheap) Parallel Computers are Here:  
**What are the Implications?**
- **Advances in Electro-optics, Imaging and LADAR Technologies**
  - MEMs Scanning Technology Revolution
    - Ultra-High Frequency Scanning (15,000 HZ ) Now Feasible
      - => **Sub-ten-picosecond light pulse time of flight detection**
      - => **As a consequence, MEMs Imaging LIDARs and LADARs appears feasible at rates of 12 million range measurements/sec swept over 30deg FOV**
        - => **HD6D scanning LADAR point clouds @ video rates ( a game changer)**
- **Advances in Machine Vision Algorithms**
  - **Novel Thread 1:** Designing *Novel Vision Sensing Hardware* to Inherently Optimize Parallel Computation and Solve The Correspondence Problem
  - **Novel Thread 2:** In the event that overlapping high density scanning lidar point clouds are measurable at video rates, we can use Thread 1 to map all pts into common coordinate system .... and couple with “statistical inference” algorithms developed to reject spurious data, characterize accuracy on the fly, and..., enable:
    - => **Super-Resolution and/or Controlled Resolution of 3D geometry, learned recursively**  
**(another game changer)**



HD scanning LADAR, to update “HD Frames”  $\sim 10^6$  meas @  $\sim 12\text{Hz}$

### Fusion of State of The Art:

- Laser pulse generation
- Multi-Channel Photo-Detection Optics and Photonics (Use “Laser Comm” SOA Photonics & Detectors)
- Novel High Speed, Multi-Beam Risley Scanner Design

### Range Measurements:

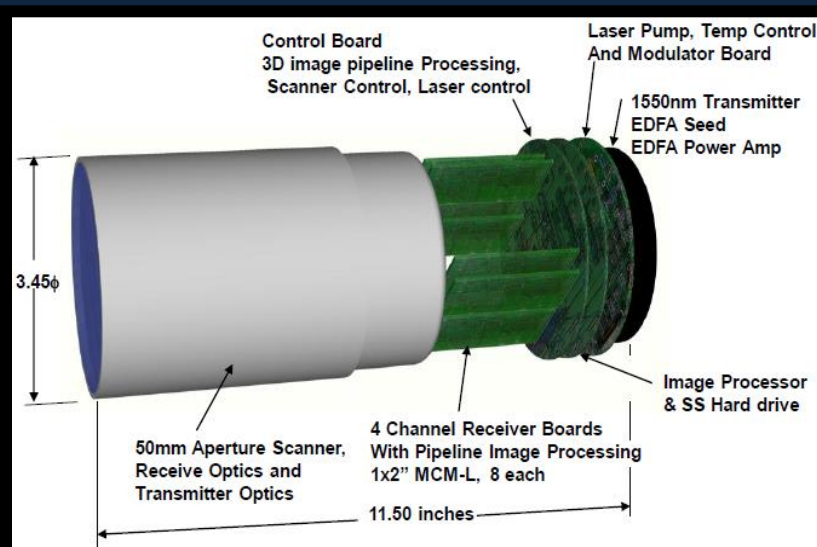
- $\pm 3\text{mm } 3\sigma$
- $\sim 12\text{MHz}$
- $30^\circ$  FOV
- Range: from 1m to 25km
- Accuracy Varies:  
 $\pm 3\text{mm}$  inside 300m range down to  $\pm 1\text{m}$  at 25km range

### Embedded High Performance Parallel Processing:

- Real-time Send/Receive Laser Pulse Waveform Convolution

### Output “Frames” of HD dense range point clouds @ 12Hz:

- Use scanner resolver measurements and IMU to compensate motion



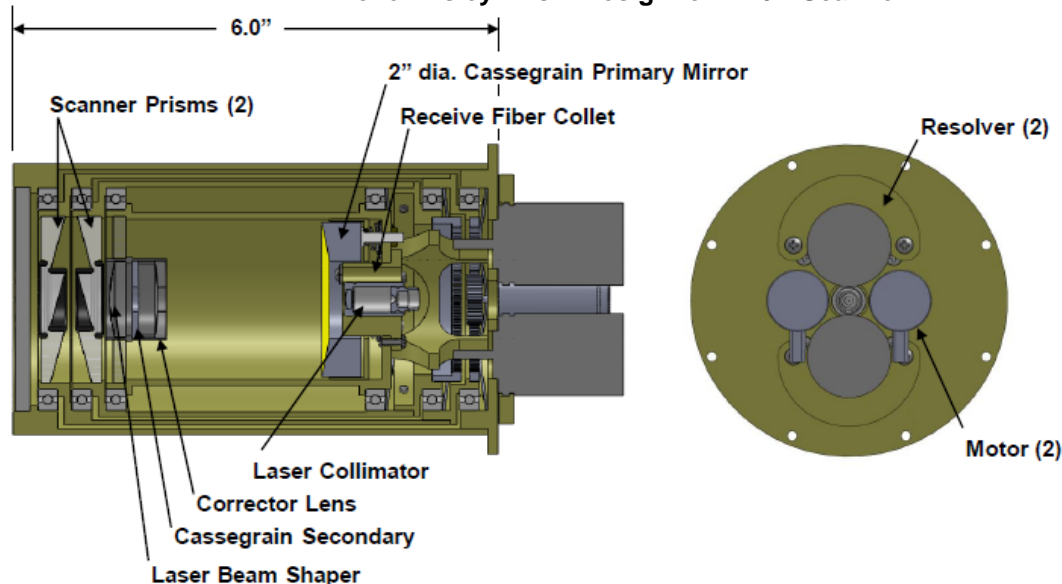
### SWAP Requirements:

- $< 2\text{kg}$
- $< 30\text{w}$
- $< 4" \times 4" \times 12"$
- Eye Safe
- 1550nm

### Prototype Delivery:

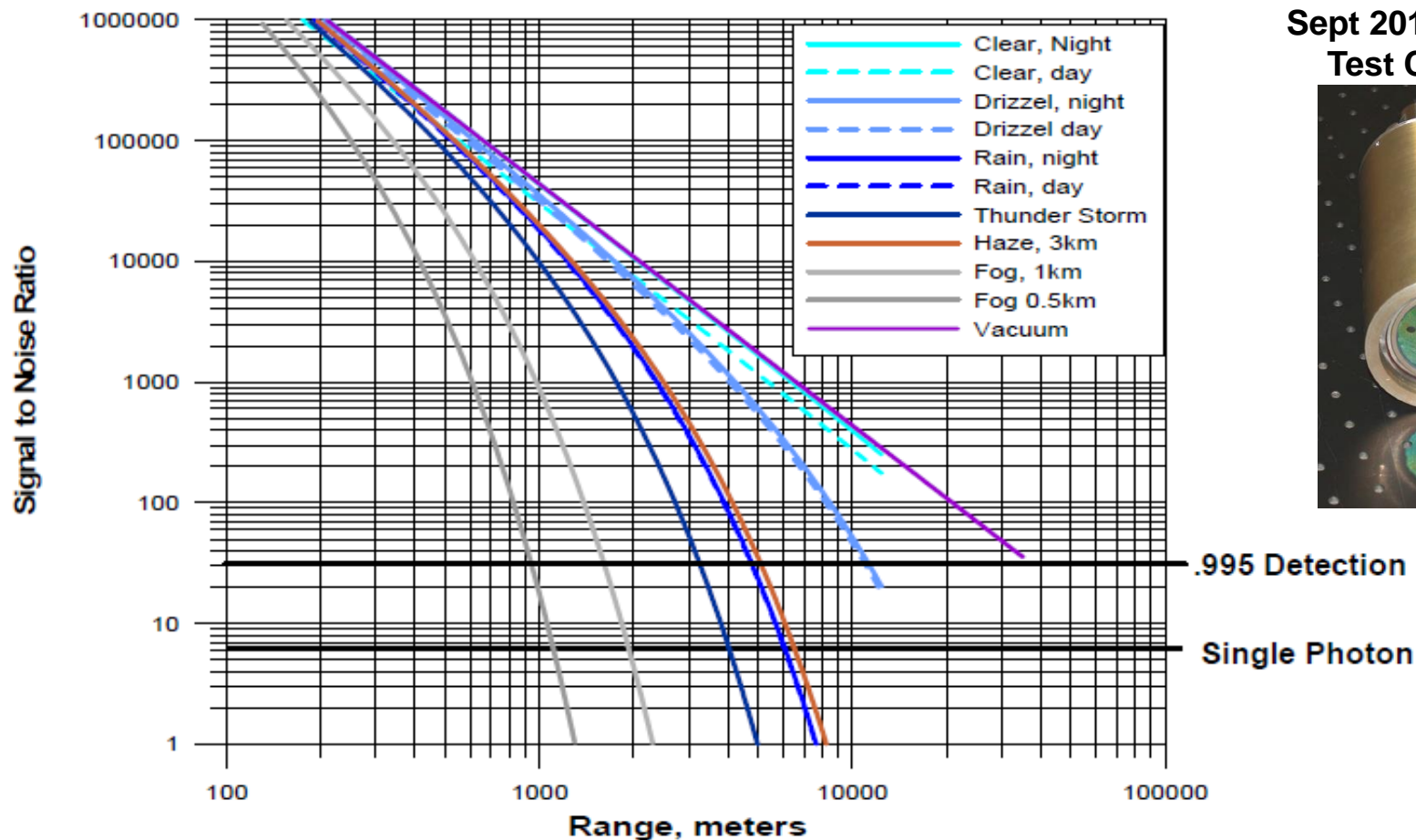
By 7/31/2011

### Novel Risley Prism Design for HD6D Scanner

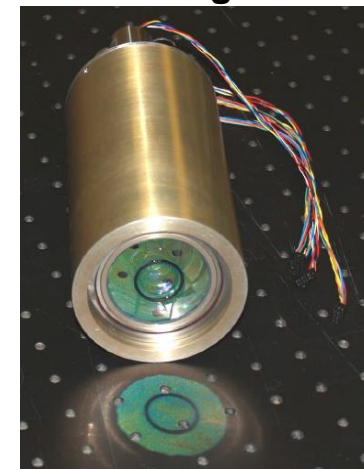


# Gen IV LADAR (HD6D): S/N vs Range

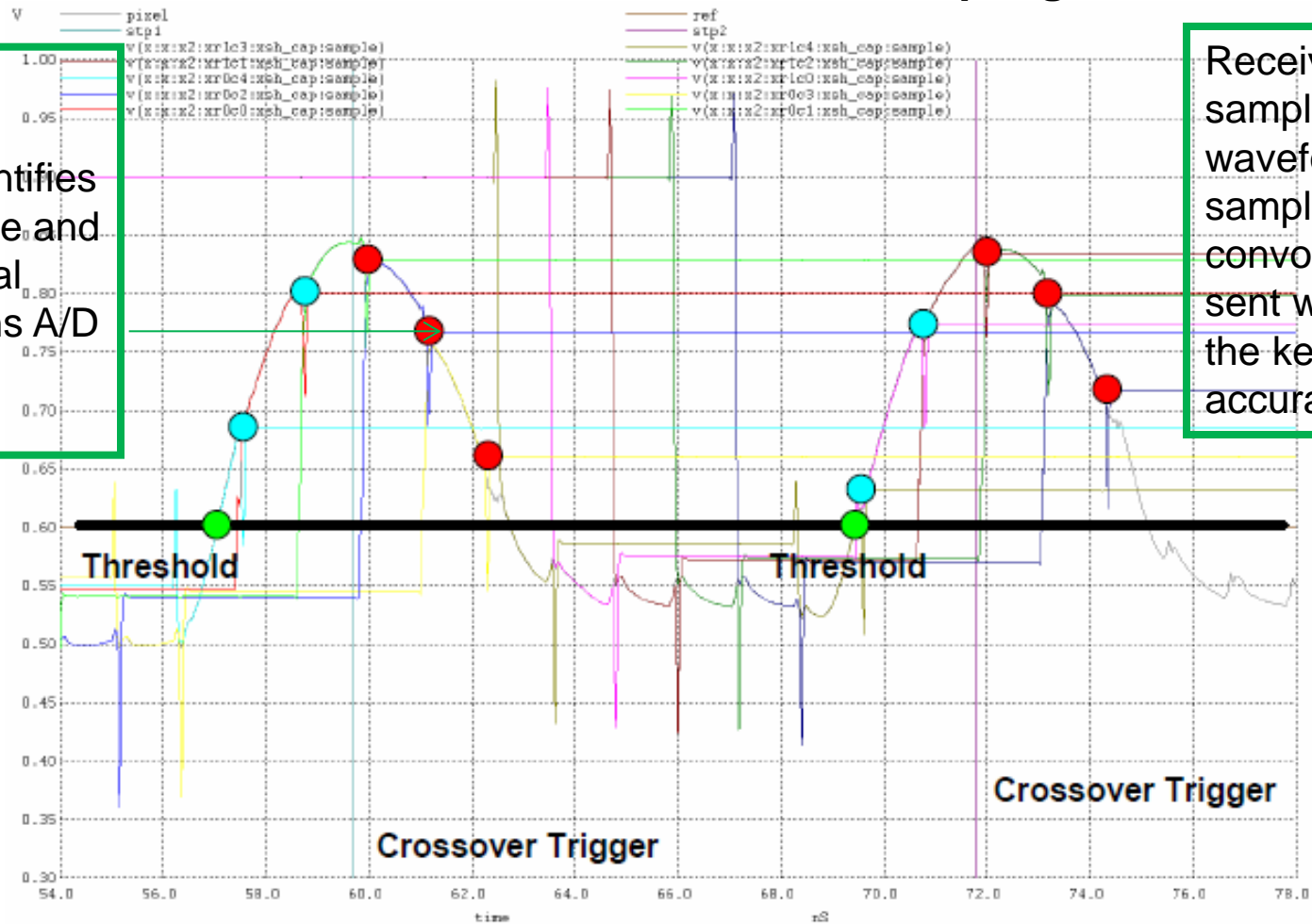
0.1 Reflectivity Target, Perpendicular to line of sight



Sept 2011 HD6D Bench  
Test Configuration



## HD6D Bench Test Waveform Sampling

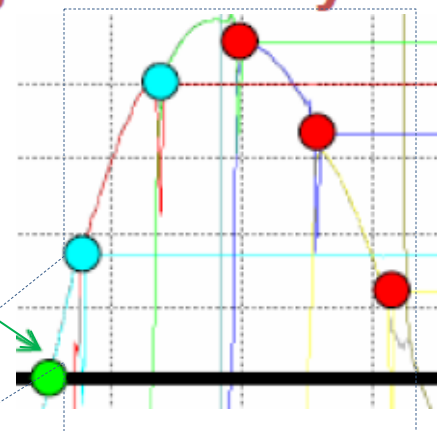
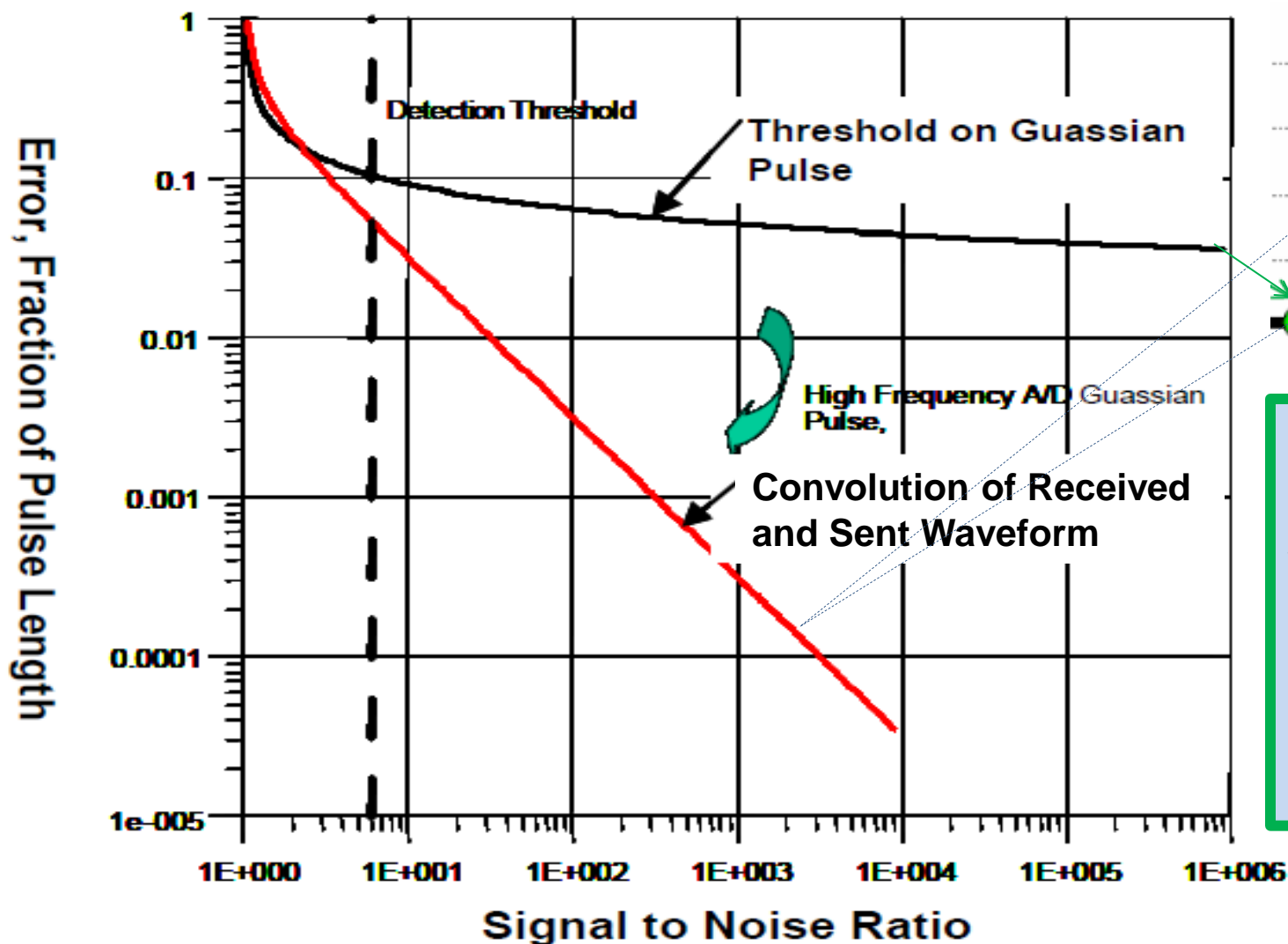


Received  
sampling of  
waveform  
sampling and  
convolution with  
sent waveform is  
the key to 3.3mm  
accuracy

● Samples before peak

● Samples after peak

## Phase Detection Advantage, Higher Range Accuracy



High precision  
Clock and Multiple  
High Precision A/D  
Waveform  
Measurements &  
Convolution Enable  
Orders of  
Magnitude Accuracy  
Improvement.



## Most Fundamental Aspects of Approach:

- TAMU/SPEC teamed to research an Ultra-Fast Scanning LADAR ...  
 → *Co-Register LADAR with a Synchronized HD Camera via Calibration*  
 => HD6D: both HD video *and* full frame HD LADAR refreshed at ~12 HZ. ...  
 => Co-optimize with algorithms and computer architecture to enable real time fusion of camera/video images with LADAR => HD6D ... => revolution!

## Research and Development Path:

- Simultaneously Refine the MEMs LADAR and the Software Pipeline
  - Waveform Processing & Algorithm Optimization=>3mm range accuracy goal
  - Develop HP\_Ransac algorithm for point cloud fusion and validation
  - Develop/implement algorithms via Pipeline & FPGA computation
    - Outcome: → *TRL 5 Prototype of the embedded system by 2<sup>nd</sup> Q 2012*
  - Opportunities for TDM → *TRL 7+ during CY 2012, 2013*

- A“Pipeline” of Novel Technologies has been conceived for future products ...
  - Synthetic Aperture Optical LADAR => HD6D from long range
  - Raman Spectral Surface Material Property Characterization
  - Real-time Holographic “Video” Visualization of HD6D Imaged Objects
  - ...

